Chaos, Solitons and Fractals 110 (2018) 41–54

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

journal homepage: www.elsevier.com/locate/chaos

Frontiers

A new centrality measure based on the negative and positive effects of clustering coefficient for identifying influential spreaders in complex networks

Kamal Berahmand, Asgarali Bouyer*, Negin Samadi

Department of Information Technology and Communications, Azarbaijan Shahid Madani University, Tabriz, Iran

ARTICLE INFO

Article history:
Received 12 October 2017
Revised 14 January 2018
Accepted 7 March 2018

Keywords:
Complex network
Node spreading
Semi-local centrality
Clustering coefficient

ABSTRACT

Identifying the most influential spreaders with the aim of reaching a maximum spreading ability has been a challenging and crucial topic so far. Many centrality measures have been proposed to identify the importance of nodes in spreader detection process. Centrality measures are used to rank the spreading power of nodes. These centralities belong to either local, semi-local, or global category. Local centralities have accuracy problem and global measures need a higher time complexity that are inefficient for large-scale networks. In contrast, semi-local measures are popular methods that have high accuracy and near-linear time complexity. In this paper, we have proposed a new semi-local and free-parameter centrality measure by applying the natural characteristics of complex networks. The proposed centrality can assign higher ranks for structural holes as better spreaders in the network. It uses the positive effects of second-level neighbors' clustering coefficient and negative effects of node's clustering coefficient in defining the importance of nodes. Therefore, the proposed centrality avoids selection of spreaders that are too close to one another. We compare the proposed method with different centrality measures based on Susceptible–Infected–Recovered (SIR) and Susceptible–Infected (SI) models on both artificial and real-world networks. Experiments on both artificial and real networks show that our method has its competitive advantages over the other compared centralities.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

A wide range of real-world phenomena, from social and information to technological and biological networks, can be described using complex networks [1–4]. Because of their shared characteristics [5], complex networks can be modeled using graphs [6]; this makes their performance analysis easy. Some of the shared structural characteristics of complex networks are power-law degree distribution, high clustering coefficient, six separation degrees, sparseness on global level, having an associative structure because of local density, etc. [7–11].

Diffusion [12–14], as a means of studying a complex network's dynamic behaviors, has been one of the most important topics in this area. Diffusion on the network is transferred from one node to another and it starts on a small scale and then affects more neighbors. In diffusion, the goal is to find influential nodes which have a higher diffusion power in comparison with other nodes. Diffusion in complex networks has a lot of applications and based on the nature of a problem, specific influential nodes can be used to accelerate, control or prevent diffusion. For instance, in marketing on social networks, the highest amount of ads can be diffused with the least amount of time and resources using influential nodes [15–19]. In computer networks, the spreading of viruses can be prevented by securing the most suitable nodes [20,21]. Another example is vaccination of people to prevent the spread of a disease [22–24]. Also, in protein and brain networks, it can be used to identify the key nodes involved in important biological applications with the goal of diagnosing a disease or developing medicine for illnesses [25–27].

Since finding influential nodes is a NP-hard problem [28], some approximate methods called centrality measures are used. Among centrality measures developed in recent years, some are more popular such as degree, betweenness, closeness, K-shell, etc. Generally speaking, these measures can be divided into three types [3]: local, semi-local, and global. Local measures only use the information from first-degree neighbors to determine the importance of a node. Therefore, they have low accuracy and time complexity. In contrast, global measures need the entire graph's information to do so and therefore have a higher accuracy and time complexity. In recent years, a new classification of measures called the semi-local

* Corresponding author.
E-mail addresses: berahmand@azaruniv.edu (K. Berahmand), a.bouyer@azaruniv.edu (A. Bouyer), n.samadi@azaruniv.edu (N. Samadi).

https://doi.org/10.1016/j.chaos.2018.03.014
0960-0779/© 2018 Elsevier Ltd. All rights reserved.

measures has been developed which have high accuracy by using more comprehensive information compared to local measures and are able to maintain an almost linear time complexity.

In the current study, it has been attempted to determine the influential nodes using the natural characteristics of complex networks in a semi-local approach. Clustering coefficient is one of the shared structural characteristics of complex networks which is studied at both micro (single node) and macro (the entire graph) levels [29]. On macro level, clustering coefficient indicates the tendency of a network's nodes to form three-member clusters. In most complex networks, the amount of clustering coefficient on global level is high. Nevertheless, having a high clustering coefficient does not necessarily mean a high clustering coefficient for all the nodes of a network. On micro level, a node's clustering coefficient is defined as the tendency of that node's neighbors to form connections with one another. In this study, our aim is to use the positive effects of second-level neighbors' clustering coefficient and negative effects of the node's clustering coefficient in defining the importance of nodes. In other words, the sum of second-level neighbors' clustering coefficient of a node indicates that the second-level of neighbors are in dense part of the graph. In addition, if that node has low clustering, high degree and dense second-degree neighboring, it is called a structural hole. In this paper, we propose a semi-local and parameter-free centrality to discover these structural hole nodes. To evaluate the efficiency of the proposed methodology, we use the Susceptible–Infected–Recovered (SIR) and Susceptible–Infected (SI) models to simulate the epidemic spreading process on both artificial and real-world networks. For our tests, we apply two SIR and SI measures to evaluate the diffusion power of top-L nodes. Moreover, we examine the ability of some centrality measures from different categories to distinguish the spreading ability of the nodes and show that our proposed method performs better. Consequently, the degree centrality from local measures, LC and LSC centralities from semi-local measures, and betweenness, closeness, and K-shell from global measures are chosen for comparison. According to the experimental results, our proposed centrality has a better performance compared to other ones and is able to have high accuracy and maintain a linear complexity at the same time.

The remainder of this paper is organized as follows: Section 2 reviews the previous related studies. The proposed semi-local centrality is discussed in Section 3. In Section 4, we present the datasets, the spreading models and the evaluation methodologies that are used to evaluate the performance of our method. Finally, Section 5 concludes the paper and gives a discussion.

2. Related work

One of the most important issues in studying the phenomenon of diffusion is finding the influential nodes [30]. The idea is to start from the right node to have a better, faster, and wider diffusion, or stop and control the negative types of diffusion through securing the most suitable nodes. To this end, many centrality measures have been offered throughout the past few years [3]. As discussed in Part 1, generally speaking, centrality measures can be divided into three types: local, semi-local, and global. Local measures only use the information from first-degree neighbors to determine the importance of a node. Since they need limited information, these measures can be used for large-scale networks regardless of their simple structure. However, they have low accuracy because of their restricted information. One of the popular measures in this category is the degree centrality. In contrast, global measures need the entire graph's information to do so, and therefore have a higher accuracy and time complexity. As a result, they cannot be used for large-scale networks. Some of the most popular measures in this category are betweenness, closeness, and K-shell centralities. In recent years, a new classification of measures called the semi-local measures has been developed to create a balance between accuracy and complexity. In semi-local measures, the analysis range is increased up to the second degree; in other words, for determining a node's importance, in addition to information from first-degree nodes, the information from second-degree nodes are used as well to increase the accuracy. This is done without increasing the time complexity of the algorithm very considerably. Some of the most popular centralities in this category are LC and LSC. On the other hand, since a graph's information may not be available all the time, measures that need more local information have received more attention. Today, the challenge is to find the influential node which also has high accuracy and low time complexity at the same time.

In this part, some of the most important centralities of each category will be analyzed. The first and simplest centrality measure in the category of local measures is the degree centrality. In this measure, only the first-degree neighbors of a node are considered important. In fact, a node is regarded as important if it has a higher degree. This measure can determine a node's importance to some extent but nodes with the same degree do not necessarily have the same essential role in the graph. This measure because of being local and ignoring the graph's global information, and with a linear time complexity of O (n), does not have high accuracy. If we have an unweight and undirect graph G = (V, E) with n = |V| nodes and m = |E| edges which has an adjacency matrix of A, degree centrality [31] is defined as Eq. (1):

$$c_D(v) = \sum_{u=1}^{n} a_{uv} = |\Gamma_1(v)| \tag{1}$$

In which Γ_1 (v) is all the first-degree neighbors of the node v.

One of the important measures in the category of global measures is the betweenness centrality. The goal of betweenness centrality is to determine the importance of a node based on the information flow existing within the graph. It is based on the number of times a node is located in the shortest paths among all the pairs of nodes in the graph. The high betweenness centrality of a node indicates that it is located between most of the shortest paths available in the graph. The betweenness centrality [32] can be defined as Eq. (2):

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{2}$$

Where σ_{st} is the number of all shortest possible paths between the pair of nodes s and t and σ_{st} (v) is the number of shortest paths between the pair of nodes s and t which also have the node v in between. Another important measure in this category is the closeness measure. This measure is based on the distance between the target node and other nodes. In this measure, the distance is calculated using the shortest paths. Therefore, a high centrality indicates that the node is located in the middle of the graph. The closeness centrality [31] can be defined as Eq. (3):

$$C_C(v) = \frac{1}{\sum_{u \in V/v} d_{uv}} \tag{3}$$

In which d_{uv} shows the shortest distance between the pair of nodes u and v.

Betweenness and closeness centralities both have a time complexity of O(n^3). These two centralities have a time complexity problem because of using the information from the shortest paths. Also, these measures limit the paths to the shortest ones while in real life, connections such as spreading of rumors, news, etc. do not necessarily happen through the shortest paths.

In recent years, a new global centrality measure called the K-shell centrality [33] has been introduced. Its decomposition

Download English Version:

https://daneshyari.com/en/article/8253720

Download Persian Version:

https://daneshyari.com/article/8253720

Daneshyari.com